

WHAT IS CLAIMED IS:

1. A system, comprising:

a psycho-physical state detection mechanism for detecting psycho-physical state of a user based on the speech from the user; and

5 a spoken dialogue mechanism for carrying on a dialogue with said user based on the psycho-physical state of the user, detected by the psycho-physical detection mechanism from the speech from the user.

10 2. The system according to claim 1, wherein said spoken dialogue mechanism comprises:

a speech understanding mechanism for understanding the speech from the user based on the psycho-physical state of the user to generate a literal meaning of the speech; and

15 a voice response generation mechanism for generating a voice response to the user based on the literal meaning of the speech and the psycho-physical state of the user.

3. The system according to claim 2, wherein said speech understanding mechanism comprises:

20 at least one acoustic model for characterizing the acoustic properties of speech, each of said at least one acoustic model corresponding to some distinct characteristic related to a psycho-physical state of a speaker;

an acoustic model selection mechanism for selecting an acoustic model that is appropriate to according to the psycho-physical state detected by the psycho-physical state detection mechanism;

a speech recognizer for generating a transcription of spoken words recognized from the speech using the acoustic model selected by the acoustic model selection mechanism; and a language understanding mechanism for interpreting the literal meaning of the speech based on the transcription.

5

4. The system according to claim 2, wherein said voice response generation mechanism comprises:

10 a natural language response generator for generating a response based on an understanding of the transcription, said response being generated appropriately according to the psycho-physical state of the user;

15 a prosodic pattern determining mechanism for determining the prosodic pattern to be applied to said response that is considered as appropriate according to the psycho-physical state; and

15 a text-to-speech engine for synthesizing the voice response based on said response and said prosodic pattern.

5. The system according to claim 1, wherein said psycho-physical state detection mechanism comprises:

20 an acoustic feature extractor for extracting acoustic features from input speech data to generate at least one acoustic feature; and

a psycho-physical state classifier for classifying the input speech data into one or more psycho-physical states based on said at least one acoustic feature.

6. The system according to claim 5, further comprising:  
at least one psycho-physical state model, each of said at least one psycho-physical state model corresponding to a single psycho-physical state and characterizing the acoustic properties of the single psycho-physical state; and

5 an off-line training mechanism for establishing said at least one psycho-physical model based on labeled training speech data.

7. The system according to claim 1, further comprising a dialogue manager that  
control the dialogue flow.

10 8. A voice based information retrieval system, comprising:  
an information database for archive information, said information being accessible and  
retrievable;  
a search engine for accessing and retrieving said information stored in the information  
15 database; and

20 a psycho-physical state sensitive spoken dialogue system connecting to the search  
engine and a user, voice communicating with the user in a psycho-physical state sensitive  
manner, responding to the user's request for desired information by activating the search  
engine to retrieve the desired information, and generating a voice response to the user  
according to the desired information and the detected psycho-physical state of the user.

9. The system according to claim 8, wherein said information database includes at least one domain information database, each of the domain information database storing the information related to at least one specific domain of interest.

5           10. A method, comprising:  
                  receiving, by a psycho-physical state detection mechanism, input speech data from a user;  
                  detecting the psycho-physical state of the user from the input speech data;  
                  understanding, by a speech understanding mechanism, the literal meaning of spoken words recognized from the input speech data based on the psycho-physical state of the user, detected by said detecting; and  
                  generating, by a voice response generation mechanism, a voice response to the user based on the literal meaning of the input speech data and the psycho-physical state of the user.

15           11. The method according to claim 10, wherein said detecting comprises:  
                  extracting, by a acoustic feature extractor, at least one acoustic feature from the input speech data; and  
                  classifying, by a psycho-physical state classifier and based on said at least one feature, the input speech data into the psycho-physical state according to at least one psycho-physical  
20           state model.

12. The method according to claim 11, further comprising:

receiving, by an off-line training mechanism, labeled training data, wherein each of the data items in said labeled training data is labeled by a psycho-physical state; and building said at least one psycho-physical state model using the labeled training data, each of the at least one psycho-physical state model corresponding to a single psycho-physical state and being established based on the data items in the labeled training data that have a label corresponding to the single psycho-physical state.

13. The method according to claim 10, wherein said understanding comprises:  
selecting, by an acoustic model selection mechanism, an acoustic model, from at least one acoustic model, that is appropriate to according to the psycho-physical state, detected by said detecting, each of said at least one acoustic model corresponding to some distinct speech characteristic related to a psycho-physical state;  
recognizing, by a speech recognizer, the spoken words from the input speech data using the acoustic model, selected by said selecting, to generate a transcription; and  
15 interpreting, by a language understanding mechanism, the literal meaning of the spoken words based on the transcription.

14. The method according to claim 10, wherein said generating comprises:  
constructing, by a natural language response generator, a natural language response 20 based on an understanding of the transcription, said natural language response being constructed appropriately according to the psycho-physical state of the user;

determining, by a prosodic pattern determining mechanism, the prosodic pattern to be applied to said natural language response, wherein the prosodic pattern is considered to be appropriate according to the psycho-physical state; and

synthesizing, by a text-to-speech engine, the voice response based on said natural language response and said prosodic pattern.

15. A method for voice based information retrieval, comprising:

communicating between a psycho-physical state sensitive spoken dialogue system and a user via voice to understand the user's request for desired information, wherein said understand is achieved according to the psycho-physical state of the user;

retrieving, by a search engine, information from an information database based on the understanding of the user's request for desired information to generate retrieved information; and

generating, by the psycho-physical state sensitive spoken dialogue system, a voice response to the user's request based on the retrieved information and the psycho-physical state of the user.

16. The method according to claim 15, wherein said communicating comprises:

receiving input speech data from the user;

detecting the psycho-physical state of the user from the input speech data; and

recognizing the user's request based on the psycho-physical state of the user.

17. The method according to claim 15, wherein said desired information includes information about at least one of:

weather;  
restaurants,  
news;  
sports;  
movies;  
stocks; and  
driving directions.

18. A computer-readable medium encoded with a program, said program comprising:  
receiving, by a psycho-physical state detection mechanism, input speech data from a  
user;  
detecting the psycho-physical state of the user from the input speech data;  
understanding, by a speech understanding mechanism, the literal meaning of spoken  
words recognized from the input speech data based on the psycho-physical state of the user,  
detected by said detecting; and  
generating, by a voice response generation mechanism, a voice response to the user  
based on the literal meaning of the input speech data and the psycho-physical state of the user.

19. The medium according to claim 18, wherein said detecting comprises:  
extracting, by a acoustic feature extractor, at least one acoustic feature from the input  
speech data; and

classifying, by a psycho-physical state classifier and based on said at least one feature, the input speech data into the psycho-physical state according to at least one psycho-physical state model.

5 20. The medium according to claim 19, further comprising:  
receiving, by an off-line training mechanism, labeled training data, wherein each of the data items in said labeled training data is labeled by a psycho-physical state; and  
building said at least one psycho-physical state model using the labeled training data, each of the at least one psycho-physical state model corresponding to a single psycho-physical state and being established based on the data items in the labeled training data that have a label corresponding to the single psycho-physical state.

10 21. The medium according to claim 18, wherein said understanding comprises:  
selecting, by an acoustic model selection mechanism, an acoustic model, from at least one acoustic model, that is appropriate to according to the psycho-physical state, detected by said detecting, each of said at least one acoustic model corresponding to some distinct speech characteristic related to a psycho-physical state;  
recognizing, by a speech recognizer, the spoken words from the input speech data using the acoustic model, selected by said selecting, to generate a transcription; and  
15 20 interpreting, by a language understanding mechanism, the literal meaning of the spoken words based on the transcription.

22. The medium according to claim 18, wherein said generating comprises:

constructing, by a natural language response generator, a natural language response based on an understanding of the transcription, said natural language response being constructed appropriately according to the psycho-physical state of the user;

5 determining, by a prosodic pattern determining mechanism, the prosodic pattern to be applied to said natural language response, wherein the prosodic pattern is considered to be appropriate according to the psycho-physical state; and

synthesizing, by a text-to-speech engine, the voice response based on said natural language response and said prosodic pattern.

10 23. A computer-readable medium encoded with a program for voice based information retrieval, said program comprising:

communicating between a psycho-physical state sensitive spoken dialogue system and a user via voice to understand the user's request for desired information, wherein said understand is achieved according to the psycho-physical state of the user;

15 retrieving, by a search engine, information from an information database based on the understanding of the user's request for desired information to generate retrieved information; and

generating, by the psycho-physical state sensitive spoken dialogue system, a voice response to the user's request based on the retrieved information and the psycho-physical state 20 of the user.

24. The medium according to claim 23, wherein said communicating comprises:  
receiving input speech data from the user;

detecting the psycho-physical state of the user from the input speech data; and  
recognizing the user's request based on the psycho-physical state of the user.